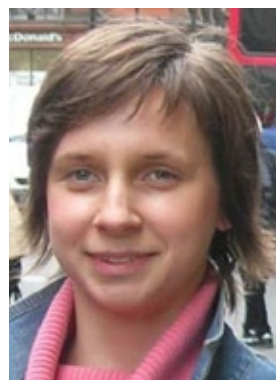




Dr Jolanta Przyłuska



Mgr Justyna Ortman
Łódź – IMP



Mgr Anna Radomska

KOLEKCJA INSTYTUTU MEDYCYNY PRACY W ŁÓDZKIEJ REGIONALNEJ BIBLIOTECIE CYFROWEJ CYBRA

Abstract

CYBRA is a digital library, which presents digitized collections of Lodz scientific libraries on one common platform. In the Nofer Institute of Occupational Medicine the following works are digitized: publications of employees of NIOM, books and journals published by the Publishing Office of the Institute as well as dissertations carried out by NIOM. Articles from selected years of journals – “*Medycyna Pracy*” and “*International Journal of Occupational Medicine and Environmental Health*” were placed on the platform as first. The article presents 2-years’ experience of NIOM obtained during development of the collection, digitization of resources (scanning along with technical processing), entering and describing records as well as contents presentation.

Streszczenie

CYBRA to biblioteka cyfrowa prezentująca na wspólnej platformie zdigitalizowane kolekcje łódzkich bibliotek naukowych. W Instytucie Medycyny Pracy im. prof. dra med. Jerzego Nofera w Łodzi (IMP) cyfryzacji poddawane są: publikacje pracowników, książki i czasopisma wydane przez Oficynę Instytutu oraz rozprawy doktorskie przeprowadzone przez Instytut. Jako pierwsze do CYBRY trafiły artykuły z wybranych roczników czasopism „*Medycyna Pracy*” oraz „*International Journal of Occupational Medicine and Environmental Health*”. Artykuł przedstawia dwuletnie doświadczenia IMP zdobyte podczas tworzenia kolekcji, digitalizacji zbiorów (skanowania wraz z obróbką techniczną), wprowadzenia i opisu rekordów oraz prezentacji treści.

Cel budowania biblioteki cyfrowej IMP

Inicjatywie budowania zasobów cyfrowych przez Bibliotekę Instytutu Medycyny Pracy przyświecały dwa podstawowe cele – pierwszy to spełnienie oczekiwań użyt-

kowników, którymi są specjaliści medycyny pracy, pracownicy stacji sanitarno-epidemiologicznych, służb bezpieczeństwa i higieny pracy, osoby podnoszące kwalifikacje zawodowe, doktoranci i studenci różnych uczelni. Użytkownicy ci są rozproszeni w różnych miejscach w kraju, a dostęp do polskich zasobów elektronicznych w tej dziedzinie jest niewielki. Drugi cel wiąże się z aspektem poznawczym tworzenia dostępu do wiedzy z wykorzystaniem rozwijających się technologii informatycznych.

Podjęciu decyzji o założeniu własnej kolekcji cyfrowej sprzyjał fakt przynależności Biblioteki IMP do Łódzkiej Akademickiej Sieci Bibliotecznej (ŁASB), gdzie jednym z wielu realizowanych zadań bibliotecznych jest budowa Łódzkiej Regionalnej Biblioteki Cyfrowej. Dodatkowym faktem umożliwiającym skanowanie dokumentów cyfrowych było utworzenie w IMP pracowni wyposażonej w profesjonalny skaner dzielowy Bookeye 3-R2 do digitalizacji piśmiennictwa w ramach projektu *Praca, Zdrowie, Środowisko – platforma informatyczna do efektywnego zarządzania wiedzą i badaniami naukowymi* z Programu Operacyjnego Innowacyjna Gospodarka. Decyzję o rozwijaniu kolekcji cyfrowej przyspieszyło dofinansowanie MNiSW na digitalizację i upowszechnianie dwóch czasopism instytutowych w ramach projektu Index Plus 2012. U podstaw tego projektu leży zapewnienie i popularyzacja cyfrowych zasobów w społeczeństwie, zwłaszcza w środowisku naukowym.

Co to jest CYBRA?

Pierwszym zbiorem dokumentów cyfrowych w regionie łódzkim była utworzona w 2005 r. biblioteka Politechniki Łódzkiej o nazwie eBiPol. Jej zasoby zostały przeniesione do biblioteki cyfrowej CYBRA wraz z jej powstaniem w 2012 r. stanowiąc pierwsze rekordy wspólnego projektu łódzkich jednostek naukowych. Obecnie CYBRA (<http://cybra.lodz.pl>) udostępnia zasoby cyfrowe pochodzące ze zbiorów następujących bibliotek łódzkich uczelni i instytucji: Politechniki Łódzkiej, Uniwersytetu Medycznego, Instytutu Medycyny Pracy, Akademii Muzycznej, Państwowej Wyższej Szkoły Filmowej, Telewizyjnej i Teatralnej, Wyższego Seminarium Duchownego, Akademii Sztuk Pięknych i Centrum Badań Molekularnych i Makromolekularnych PAN.

Dostęp do jej zbiorów jest bezpłatny. Każdy obiekt cyfrowy można przeglądać w pełnym tekście oraz pobierać na swój komputer - w zakresie, na jaki pozwala licencja udzielona przez autora lub obowiązujące przepisy prawa autorskiego. Ze zbiorów mogą korzystać wszyscy, chyba że w opisie danego obiektu cyfrowego znajduje się informacja, że dokument jest dostępny tylko w sieci wewnętrznej danej instytucji w zakresie dozwolonego użytku¹.

¹ O projekcie CYBRA. Dostęp: <http://cybra.lodz.pl/dlibra/text?id=example-page> [odczyt: 17 VII 2014]

Oprogramowanie

Obiekty cyfrowe udostępniane są w CYBRZE dzięki oprogramowaniu dLibra. Jest to pierwszy polski system do budowy bibliotek cyfrowych tworzony przez Poznańskie Centrum Superkomputerowo-Sieciowe od 1999 r. dLibra jest szeroko wykorzystywana przez instytucje, takie jak biblioteki, muzea czy archiwa do udostępniania cyfrowych postaci gromadzonych przez nie zbiorów. Obecnie jest to najpopularniejszy system do budowy bibliotek cyfrowych w Polsce. Do dnia 8 sierpnia 2014 r. system wdrożono w 85 jednostkach, ich pełna lista znajduje się pod adresem <http://dlibra.psnc.pl/dlibra/deployments/>. Oprócz ogólnie dostępnych instalacji, dLibra jest wykorzystywana również do zastosowań wewnętrznych, bez możliwości publicznego dostępu przez Internet. Są to zazwyczaj wdrożenia przeznaczone do celów dydaktycznych, jak i na potrzeby wewnętrzne danej instytucji (np. Akademii Obrony Narodowej, Urzędu Miasta Torunia czy Instytutu Pamięci Narodowej)².

System dLibra posiada szereg funkcji takich jak:

- możliwość przechowywania obiektów cyfrowych w dowolnym formacie, np. PDF, DjVu, MP3, FLV, JPG,
- dedykowane narzędzia wspomagające importowanie danych z zewnętrznych systemów informatycznych,
- zaawansowane możliwości zarządzania przechowywaną treścią oraz metadanymi,
- bezpieczne i elastyczne mechanizmy uwierzytelniania,
- wsparcie dla budowania rozszerzonych bibliotek cyfrowych (np. regionalnych), zrzeszających setki instytucji³.

CYBRA wraz z innymi ogólnodostępnymi bibliotekami tworzy platformę rozproszonych polskich bibliotek cyfrowych w sieci PIONIER, dając dostęp do setek tysięcy obiektów cyfrowych i ich metadanych oraz do zaawansowanych usług opartych na wspólnych obiektach, takich jak wyszukiwanie rozproszone lub profil użytkownika sieci. Integrujący element platformy jest dostępny od czerwca 2007 r. w postaci Federacji Bibliotek Cyfrowych (<http://fbc.pionier.net.pl/>).

Kolekcja Instytutu Medycyny Pracy

Zasoby

Zarządzanie zasobami cyfrowymi IMP odbywa się poprzez Łódzką Akademicką Sieć Biblioteczną (ŁASB), która współpracuje z Zakładem Akademickiej Sieci Bibliotecznej (ZASB) przy Centrum Komputerowym Politechniki Łódzkiej. Infrastrukturę

² Władysław Marek Kołasa: dLibra Digital Framework – platform do budowy bibliotek cyfrowych. W: Biblioteki Cyfrowe: projekty, realizacje, technologie. Warszawa: Wydawnictwo SBP 2007 s. 67-88

³ dLibra: Biblioteki cyfrowe. Dostęp: <http://dingo.psnc.pl/dlibra/> [odczyt: 28 VII 2014]

sieciowo-sprzętową dostarcza Miejska Sieć Komputerowa LODMAN. ŁASB realizuje wspólnie uzgodnione przedsięwzięcia i projekty, w tym tworzenie regionalnej biblioteki cyfrowej. Jest to usługa rozszerzona dla tych bibliotek, które zdecydowały się na pokazanie swoich zasobów na wspólnej platformie CYBRY. Administrowanie systemem, upgrade oprogramowania, kopie bezpieczeństwa, archiwizowanie zapewniane są w ramach umowy z ŁASB. Po stronie pracowników Działu Zarządzania Wiedzą IMP leży przygotowanie, opracowanie i wprowadzenie dokumentów cyfrowych. Pierwsze prace nad budową kolekcji cyfrowej Instytutu Medycyny Pracy w bibliotece CYBRA rozpoczęto w 2012 r. Obecnie zasoby podzielone są na kolekcje (w języku polskim i angielskim):

- książki (books)
- czasopisma (journals)
- sympozja, kongresy, konferencje (symposia, congresses, conferences)
- ludzie medycyny pracy (people of occupational medicine)
- dysertacje (dissertations)
- varia (others).



Rys. 1. Opis kolekcji „Czasopisma” zasobów IMP w Bibliotece Cyfrowej CYBRA.



Rys. 2. Grafiki zaprojektowane przez IMP dla poszczególnych kolekcji w Bibliotece Cyfrowej CYBRA.

Każda z kolekcji posiada swój unikatowy opis (rys. 1) wraz z grafiką charakterystyczną dla umieszczonych w niej zasobów (rys. 2).

Jako pierwsze do udostępniania zostały przeznaczone artykuły z czasopism wydawanych w Instytucie: „*Medycyna Pracy*” z lat 2002–2011 i „*International Journal of Occupational Medicine and Environmental Health*” z lat 1988–2010. Wybór tych lat wynika z założeń projektu Index Plus, w którym przewidziano uzupełnienie dostępu do czasopism niezapewnione dotychczas przez inne bazy i narzędzia.

Zasoby cyfrowe kolekcji IMP w CYBRZE, czyli obiekty i metadane, są utrzymywane równolegle z zasobami drukowanymi gromadzonymi w Bibliotece, stanowiąc dodatkową formę udostępniania. Dokumenty cyfrowe skierowane są do odbiorców zlokalizowanych na terenie całego kraju, a także dla użytkowników zagranicznych. Publikacje wprowadzane są równolegle w dwóch językach: polskim i angielskim, tworząc bazę artykułów o szerokim zasięgu terytorialnym.

Wszystkie zasoby opisywane są w sposób pozwalający na uzyskanie pełnej informacji o tym co zawiera kolekcja IMP, za jaki okres gromadzone są dokumenty, podany jest format ich udostępniania, prawa dostępu, typ zasobu, identyfikator czy lokalizacja oryginału.

Zasoby CYBRY rejestrowane są w bibliotekach i katalogach:

- FBC – Federacji Bibliotek Cyfrowych (<http://fbc.pionier.net.pl>),
- Europeana (<http://www.europeana.eu/>),
- ROAR – Registry of Open Access Repositories (<http://roar.eprints.org/>),
- OpenDOAR – The Directory of Open Access Repositories (<http://www.open-doar.org>).

Miarą użyteczności zasobów CYBRY są wskaźniki liczbowe pokazujące najczęściej przeglądane dokumenty, widoczne na głównej stronie kolekcji obok informacji o ostatnio dodanych publikacjach (rys. 3.).

Szczegółowe raporty użytkowania analizowane są za pomocą generatora statystyk AWStats. Statystyki odwiedzin zawierają dane na temat liczby unikatowych gości, wizyt, przeglądanych stron, historii odwiedzin miesięcznych, krajów z których przeglądano serwis, czasu wizyt, najczęściej poszukiwanych fraz tekstowych czy słów kluczowych. Dane te po dłuższym okresie użytkowania kolekcji IMP będą pomocne przy ustalaniu polityki rozwoju zasobów w CYBRZE.



Rys. 3. Najczęściej przeglądane i ostatnio dodane artykuły na stronie kolekcji IMP

Łódzka Regionalna Biblioteka Cyfrowa stopniowo poszerza swoją zawartość o dokumenty ze zbiorów kolejnych bibliotek z regionu łódzkiego. Każda kolekcja jest budowana indywidualnie. Powstaje w ten sposób zbiór dokumentów, którego spoiwem jest narzędzie do udostępniania. Dla podkreślenia różnorodności kolekcji każda instytucja stosuje własny projekt graficzny i inną kolorystykę stron. W kolekcji IMP udostępniane są dokumenty świadczące o tematyce badań z medycyny pracy i rozwoju dziedziny z przeznaczeniem ich dla użytkowników ze środowiska naukowego. Dokumenty opisywane są za pomocą metadanych wg Dublin Core Metadata Initiative zawierających elementy służące współdziałaniu z innymi systemami. Dane można wyeksportować do formatów:

- OAI (ma na celu zapewnienie trwałego źródła dla danych w Open Archives Initiative),
- RDF (opisuje zawartość grupy danych w zwięzłym pliku tekstowym w sposób zrozumiały dla programów komputerowych),
- RIS (umożliwia menadżerom cytowań wymianę danych),
- BIBTEX (ułatwia dostosowanie uprzednio zebranych danych bibliograficznych do wymogów różnych wydawców).

Obiekty

W kolekcji biblioteki cyfrowej IMP gromadzone są zarówno obiekty „born digital” oraz „digital reformatting” (wtórnie cyfrowe). Pierwsze to takie, którym w momencie tworzenia została nadana forma cyfrowa. Natomiast drugie pierwotnie występowały w formie tradycyjnej i aby trafić do zasobów kolekcji musiały przejść proces digitalizacji. Proces ten składa się m.in. ze skanowania, obróbki graficznej oraz przeprowadzenia „Optical Character Recognition” (OCR). Tak przygotowany obiekt poddawany jest procesom, jakie zazwyczaj od razu stosuje się w przypadku publikacji „born digital” – nadanie znaku wodnego, zapisanie kopii w formacie archiwizacyjnym oraz opisanie metadany.

Dzięki zastosowaniu OCR wszystkie dokumenty tekstowe kolekcji IMP można przeszukiwać pełnotekstowo i kopiować wybrane fragmenty, co znacząco rozszerza ich użyteczność.

Zasoby umieszczane w CYBRZE zapisywane są w kilku formatach zależnych od parametrów obiektu. W przypadku plików tekstowych najczęściej stosowany jest format PDF. W sytuacji, gdy plik wymaga zabezpieczeń przed próbami kopiowania lub gdy posiada bardzo duże rozmiary stosowany jest format DjVu. Dla obiektów graficznych (takich jak miniatury okładek widocznych przy każdym rekordzie) używany jest format TIFF.

Tym co odróżnia kolekcję IMP od innych jest wprowadzenie wszystkich pojedynczych artykułów czasopisma jako osobnych plików PDF. Dzięki temu każda publikacja posiada indywidualny, rozbudowany opis i ma większe szanse na wyszukanie przez potencjalnego odbiorcę. Zapewnienie trwałości obiektów zamieszczonych w CYBRZE odbywa się automatycznie poprzez cotygodniowe tworzenie ich kopii zapasowych. Tak zapisane pliki przechowywane są na dyskach magnetycznych oraz na płytach DVD archiwizowanych w bibliotece IMP.

Metadane

Przy opisie bibliograficznym Łódzka Regionalna Biblioteka Cyfrowa korzysta ze standardu opisu metadanych – Dublin Core w wersji 1.1. Oprócz podstawowych 15 atrybutów Dublin Core wykorzystuje tzw. kwalifikatory uszczegółowiające opis⁴. Przykładem atrybutu jest w tym przypadku pole: *Wydawca*, a kwalifikatorem uszczegółowiającym go: *Miejsce wydania*. Pełna struktura atrybutów i kwalifikatorów wykorzystywanych przez IMP wygląda następująco:

- Tytuł
- Autor

⁴ ePoradnik redaktora zasobów cyfrowych. Dostęp: <http://www.bibliotekacyfrowa.pl/dlibra/docmetadata?id=25069&from=publication> [24 VII 2014]

- Współtwórca
- Promotor
- Wydawca
 - Miejsce wydania
- Data wydania
- Źródło
- Prawa
 - Właściciel praw
 - Prawa dostępu
 - Licencja
- Słowa kluczowe
- Uwagi
- Język
- Format
- Typ zasobu
- Powiązania
- Zakres
- Identyfikator
- Tagi
- Digitalizacja
- Lokalizacja oryginału

Uzupełnianie treści we wszystkich polach nie jest obligatoryjne. W dużym stopniu zależy to od potrzeb danej biblioteki. Jednak, by opis był poprawnie zidentyfikowany i zaindeksowany w bibliotece cyfrowej konieczne jest zapewnienie pewnego minimum danych, w tym przypadku są to atrybuty: autor, tytuł, data publikacji.

Ze względu na stosowanie dwóch języków (polskiego i angielskiego) do opisów bibliograficznych podczas wprowadzania metadanych zaistniała potrzeba używania zakładek:

1. opisujących obiekt w języku polskim,
2. opisujących obiekt w języku angielskim,
3. opisujących metadane niezależne (stałe bez względu na wersję językową), takie jak autor czy identyfikator.

Aplikacja wyposażona jest w moduł, który podczas wprowadzania publikacji porównuje jej atrybut „Tytuł” z tymi już umieszczonymi w bazie biblioteki i informuje, jeśli dany obiekt został tam już zamieszczony. Takie rozwiązanie pozwala uniknąć dublowania tych samych publikacji.

Ważną grupą metadanych (zawsze wprowadzaną w bibliotece cyfrowej IMP) są atrybuty, wskazujące kto jest właścicielem praw autorskich oraz kto i na jakich zasa-

dach może z danego obiektu korzystać. Kolekcja IMP podzielona jest na dwie części. W pierwszej znajdują się artykuły, których autorami są pracownicy IMP. Na podstawie przysługującej Instytutowi licencji udostępniane są one wszystkim zainteresowanym użytkownikom Internetu. W drugiej grupie znajdują się obiekty, których właścicielami są autorzy spoza IMP lub inne instytucje. W ich przypadku postępuje się zgodnie z *Ustawą o prawie autorskim i prawach pokrewnych* (Art. 28 pkt.3)⁵. Oznacza to, że te publikacje w formie pełnotekstowej dostępne są wyłącznie dla użytkowników komputerów znajdujących się w Bibliotece Instytutu. Natomiast dla pozostałych użytkowników widoczny jest jedynie ich opis bibliograficzny.

W ciągu 24 godzin od dodania publikacji do zasobów Łódzkiej Regionalnej Biblioteki Cyfrowej zostaje ona zindeksowana protokołem OAI (Open Archives Initiative). W ten sposób system biblioteki, który jest z nim zgodny udostępnia metadane narzędziom indeksującym umożliwiając wykorzystywanie w innych serwisach i bazach danych⁶.

⁵ Tekst ujednolicony ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych, Art. 28 pkt. 3. Dostęp: <http://isap.sejm.gov.pl/DetailsServlet?id=WDU19940240083> [04 VIII 2014]

⁶ Marek Nahotko: *Biblioteki cyfrowych książek w środowisku akademickim i bibliotekarskim*. Warszawa: Wyd. CPI 2006 s. 177-188